

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14533 A2

(51) International Patent Classification⁷: C12N 9/00

(21) International Application Number: PCT/GB00/03100

(22) International Filing Date: 15 August 2000 (15.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9919766.7 21 August 1999 (21.08.1999) GB

(71) Applicant (for all designated States except MG, US): AS-
TRAZENECA AB [SE/SE]; S-151 85 Sodertalje (SE).

(71) Applicant (for MG only): ASTRAZENECA UK LIM-
ITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN
(GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): ROSAMOND,

John, David, Charles [GB/US]; 35 Gatehouse Drive,
Waltham, MA 02451 (US). SCHNELL, Norbert, Friede-
mann [DE/US]; 35 Gatehouse Drive, Waltham, MA 02451
(US).

(74) Agent: GAINEY, Laurence, David, Scott; Astrazeneca,
Global Intellectual Property, PO Box 272, Mereside, Alder-
ley Park, Macclesfield, Cheshire SK10 4GR (GB).

(81) Designated States (national): JP, MG, US.

(84) Designated States (regional): European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
NL, PT, SE).

Published:

— Without international search report and to be republished
upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

ATTORNEY DOCKET NUMBER: 10182-016-999
SERIAL NUMBER: 10/032,585
REFERENCE: CF

WO 01/14533 A2

(54) Title: PROTEIN

(57) Abstract: This invention relates to newly identified polynucleotides, polypeptides encoded by these polynucleotides, to the production of such polynucleotides and polypeptides, and to the uses of such polynucleotides and polypeptides. More specifically, the invention relates to the phosphomevalonate kinase (PMK) gene (ERG8 gene) from *Candida Albicans* (*C. albicans*), to methods for its expression yielding phosphomevalonate kinase protein, to novel hybrid organisms for use in such expression methods, to methods for purification of the protein, to methods and tools for diagnosing *C. albicans* infection and to assays for identifying inhibitors of the enzyme which inhibitors have potential as anti-fungal agents.

PROTEIN

This invention relates to newly identified polynucleotides, polypeptides encoded by these polynucleotides, to the production of such polynucleotides and polypeptides, and to the
5 uses of such polynucleotides and polypeptides. More specifically, the invention relates to the phosphomevalonate kinase (PMK) gene (ERG8 gene) from *Candida albicans* (*C. albicans*), to methods for its expression yielding phosphomevalonate kinase protein, to novel hybrid organisms for use in such expression methods, to methods for purification of the protein, to methods and tools for diagnosing *C. albicans* infection and to assays for identifying inhibitors
10 of the enzyme which inhibitors have potential as anti-fungal agents.

C. albicans is an important human fungal pathogen and the most prominent target organism for antifungal research. PMK is an enzyme required for the biosynthesis of isoprene subunits that are used as precursors in the synthesis of sterols, dolichols and ubiquinones. As PMK is an essential biosynthetic enzyme, inhibitors of PMK should find use as antifungal
15 agents. All species synthesise a protein with PMK activity however, across species the enzymes differ considerably in their amino acid sequence. Because of selectivity problems (for example fungal versus human) it is extremely important to optimise potential inhibitors specifically against the fungal target enzymes (i.e. *C. albicans* or *Aspergillus fumigatus*) and not against the human enzyme. Such cross-fungal-species inhibitors possess broad specificity.
20 Alternatively, it may be desirable to use an inhibitor which is more selective, for example, one that inhibits *C. albicans* PMK but not a homologous but non-identical fungal PMK protein such as from *Saccharomyces cerevisiae* (*S. cerevisiae*).

In view of the increased incidence of fungal resistance to existing anti-fungal agents and fuelled by the growing number of fungal infections particularly in people with
25 immunodeficiency disorders, organ transplants and cancer, there is a need for new means of identifying potential anti-fungal agents.

We have now successfully cloned the ERG8 gene from *C. albicans* (hereinafter referred to as ERG8 gene) and determined its full length nucleotide sequence and corresponding (PMK) polypeptide sequence (hereinafter referred to as ERG8 protein) as set
30 out in Figure 1 and SEQ ID No. 7 of this application respectively. The coding DNA sequence (SEQ ID NO. 6) of the *C. albicans* ERG8 gene isolated is 1299 nucleotides in length and the corresponding protein sequence is 433 amino acids in length (SEQ ID NO. 7). The protein exhibits approximately 45% homology with the corresponding protein from *S. cerevisiae* and

only about 10% homology to that of the human protein equivalent. Homology as used herein, takes the definition known to and routinely used by molecular biologists. It refers to the sequence identity between two sequences as assessed by best-fit computer alignment analysis using suitable software such as Blast, Blast2, NCBI Blast2, WashU Blast2, FastA, Fasta3 and
5 PILEUP, using a scoring matrix such as Blosom 62. Such software packages endeavour to closely approximate the "gold-standard" alignment algorithm of Smith-Waterman. Thus, the preferred software/search engine programme for use in assessing the percent identity or similarity, i.e how two primary polypeptide sequences line up is Smith-Waterman. Identity refers to direct matches, similarity allows for conservative substitutions.

10 According to a first aspect of the invention there is provided an isolated or purified polypeptide which is ERG8 protein, as well as variants thereof. The preferred polypeptide sequence is that as set out in SEQ ID NO. 7. The complete *C. albicans* phosphomevalonate kinase enzyme polypeptide has the amino acid sequence as depicted in SEQ ID No. 7 herein. The polypeptides of the present invention include the polypeptide of SEQ ID No. 7 as well as
15 polypeptides which have in increasing order of preference, at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% identity to the polypeptide whose amino acid sequence is depicted in SEQ ID NO. 7.

As used herein, the term "isolated" refers to molecules, either nucleic acid or amino acid sequences, that are removed from their natural environment and purified or separated
20 from at least one other component with which they are naturally associated. Also encompassed by this term are molecules that are artificially synthesised and purified away from their synthesis materials. Thus, a polynucleotide is said to be isolated when it is substantially separated from other contaminant polynucleotides or nucleotides.

Although the natural polypeptide of SEQ ID NO. 7 and a variant polypeptide may only
25 possess for example 80% identity, they are actually likely to possess a higher degree of similarity, depending on the number of dissimilar codons that are conservative changes. Similarity between two sequences includes direct matches as well as conserved amino acid substitutes which possess similar structural or chemical properties, e.g. similar charge. Examples of conservative changes (conserved amino acid substitutes) are *inter alia*: alanine to
30 glycine, isoleucine, valine or leucine; tyrosine to phenylalanine or tryptophan; and lysine to arginine or histidine.

Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made without altering the biological activity of the resulting polypeptide,

regardless of the chosen method of synthesis. The phrase "conservative substitution" includes the use of a chemically derivatized residue in place of a non-derivatized residue provided that such polypeptide displays the desired binding activity. D-isomers as well as other known derivatives may also be substituted for the naturally occurring amino acids. See, e.g., U.S.

- 5 Patent No. 5,652,369, *Amino Acid Derivatives*, issued July 29, 1997. Substitutions are preferably, although not exclusively, made in accordance with those set forth in TABLE 1 as follows:

TABLE 1

Original residue	Example conservative substitution
Ala (A)	Gly; Ser; Val; Leu; Ile; Pro
Arg (R)	Lys; His; Gln; Asn
Asn (N)	Gln; His; Lys; Arg
Asp (D)	Glu
Cys (C)	Ser
Gln (Q)	Asn
Glu (E)	Asp
Gly (G)	Ala; Pro
His (H)	Asn; Gln; Arg; Lys
Ile (I)	Leu; Val; Met; Ala; Phe
Leu (L)	Ile; Val; Met; Ala; Phe
Lys (K)	Arg; Gln; His; Asn
Met (M)	Leu; Tyr; Ile; Phe
Phe (F)	Met; Leu; Tyr; Val; Ile; Ala
Pro (P)	Ala; Gly
Ser (S)	Thr
Thr (T)	Ser
Trp (W)	Tyr; Phe
Tyr (Y)	Trp; Phe; Thr; Ser
Val (V)	Ile; Leu; Met; Phe; Ala

- 10 The nucleotide sequences of the present invention may also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which

modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, eg, site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference, etc.

- 5 Included within the scope of the present invention are alleles of the ERG8 molecule of the present invention. As used herein, an "allele" or "allelic sequence" is an alternative form of the kinase molecule described herein. Alleles result from nucleic acid mutations and mRNA splice-variants which produce polypeptides whose structure or function may or may not be altered. Any given gene may have none, one or many allelic forms. Common
- 10 mutational changes which give rise to alleles are generally ascribed to natural deletions, additions or substitutions of amino acids. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

Thus, according to a preferred embodiment there is provided an isolated polypeptide comprising the sequence depicted in SEQ ID No. 7 or a sequence possessing at least 80%

15 similarity thereto. More preferred embodiments are those that have in increasing order of preference at least 85, 90, 95, 96, 97, 98 and 99% similarity to the sequence depicted in SEQ ID No. 7. Functional biologically active variants are preferred.

- Fragments of such polypeptides comprising at least 15, preferably at least 30 and more preferably at least 50 contiguous amino acids are also encompassed by the present invention.
- 20 Such fragments may be used as intermediates to generate longer polypeptide fragments including preferably, the full-length polypeptide sequence as depicted in SEQ ID No. 7, or a functional variant thereof. Such polypeptide fragments may also be used to raise antibodies against or specific for parts of the ERG8 protein.

The invention also relates to variant polypeptide sequences encoded by nucleic acid

25 capable of hybridising with nucleic acid coding for the natural polypeptide (SEQ ID No. 6, or its complementary antisense strand)(or would do so but for the degeneracy of the genetic code), for example under stringent conditions (such as at 35°C to 65°C in a salt solution of approximately 0.9M). Such hybridisable polynucleotides are also part of the invention. The present invention particularly relates to polynucleotides which hybridise to the ERG8

30 polynucleotide sequence depicted in SEQ ID NO. 6, its complementary sequence, or fragment thereof, under stringent conditions. As used herein, stringent conditions are those conditions which enable sequences that possess at least 80%, preferably at least 90% and more preferably

at least 95% sequence identity to hybridise together. Thus, nucleic acids which can selectively hybridise to the nucleic acid of SEQ ID No. 6, or the complementary antisense strand thereof, include nucleic acids which have at least 80%, preferably at least 90%, more preferably at least 95%, still more preferably at least 98% sequence identity and most preferably 100%,
5 over at least a portion of the nucleic acid encoding the ERG8 gene disclosed herein. Selectively hybridise means that the molecule must be capable of specifically hybridising to the nucleic acid sequence of SEQ ID No. 6 or its complement, to the exclusion of other naturally occurring sequences. As well as full-length gene sequences, smaller nucleic acid fragments for example oligonucleotide primers which can be used to amplify the ERG8 gene
10 using any of the well known amplification systems such as polymerase chain reaction (PCR), or fragments that can be used as diagnostic probes to identify corresponding nucleic acid sequences are also part of this invention. The invention thus includes polynucleotides of shorter length than the full length ERG8 gene sequence depicted in SEQ ID No. 6, that are capable of specifically hybridising to the nucleic acid encoding the *C. albicans* ERG8 gene
15 described herein. Such polynucleotides may be at least 10 nucleotides in length, preferably at least 15, more preferably at least 20 and most preferably at least 30 nucleotides in length and may be of any size up to and including the full length ERG8 nucleotide sequence. The presence of mismatch nucleotides in the hybridisation polynucleotides is not detrimental to the utility of such polynucleotides provided that they are capable of selectively hybridising to the
20 target ERG8 nucleotide sequence.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe nucleic acid is greater than 500 bases or base pairs is: 6 x SSC (saline sodium citrate), 0.5% SDS (sodium dodecyl sulphate), 100µg/ml denatured, sonicated salmon sperm DNA. The hybridisation being performed at 68°C for at least 1 hour
25 and the filters then washed at 68°C in 1 x SSC, or for higher stringency, 0.1 x SSC/0.1% SDS.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe is an oligonucleotide of between 12 and 50 bases is: 3M trimethylammonium chloride (TMACl), 0.01M sodium phosphate (pH 6.8), 1mM EDTA (pH
30 7.6), 0.5% SDS, 100µg/ml denatured, sonicated salmon sperm DNA and 0.1 dried skimmed milk. The optimal hybridisation temperature (T_m) is usually chosen to be 5°C below the T_i of the hybrid chain. T_i is the irreversible melting temperature of the hybrid formed between the

probe and its target. If there are any mismatches between the probe and the target, the T_m will be lower. As a general guide, the recommended hybridisation temperature for 17-mers in 3M TMACl is 48-50°C; for 19-mers, it is 55-57°C; and for 20-mers, it is 58-66°C.

A suitable hybridisation protocol is described in Example 5 herein, however, operable variations to this method will be apparent to the person skilled in the art.

As used herein, the term 'variant' includes naturally occurring allelic variants as well as non-naturally occurring variants, fragments and analogs of the sequences depicted in SEQ ID NOs. 6 or 7. Such variants include C- or N-truncated variants, deletion variants, substitution variants as well as addition and insertion variants. The term 'analog' refers to proproteins which can be activated by cleavage of the proprotein portion to release the biologically active polypeptide or protein. The term 'derivative' refers to a polypeptide encoded by a chemically modified ERG8 gene, for example one wherein hydrogen has been replaced by an acyl or amino group, as well as polypeptides possessing one or more non-natural amino acids. When referring to a polypeptide or protein sequence, a functional variant is one that has retained at least some PMK enzymatic activity. The variant polypeptides of the present invention may comprise internal, but preferably, terminal flanking sequences (fusion proteins) to facilitate protein purification. Such 'additional domain' sequences (Flag sequences) may comprise for example, metal chelating peptides such as histidine-tryptophan modules (including 6-his tags) that allow purification of the polypeptide on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, or peptide domains that allow purification on immobilised antibodies specific for the peptide. Other suitable 'additional purification domains' will be known to the person skilled in the art.

According to a preferred embodiment of the invention the native ERG8 polypeptide sequence (having the sequence as depicted in SEQ ID No. 7) is fused at its amino terminus to six histidine residues which serve to enable the polypeptide, once expressed from the host cell, to be isolated and purified by affinity chromatography using a Ni-chelate resin.

A flanking purification domain may be separated from the ERG8 polypeptide by a cleavage sequence such as that recognised by thrombin or Factor Xa so as to facilitate release of the polypeptide from the flanking sequence which may or may not be attached to an immobilised support. Alternatively, cyanogen bromide which cleaves at methionine residues can be employed to release the desired polypeptide from its flanking sequence.

The polypeptides of the invention can be synthesised chemically. For example, by the Merryfield technique (J. Amer. Chem. Soc. 85:2149-2154, 1968). Numerous automated polypeptide synthesisers, such as Applied Biosystems 431A Peptide Synthesizer also now exist. Alternatively, and preferably, the polypeptides of the invention are produced from a
5 nucleotide sequence encoding the polypeptide using recombinant expression technology.

In a further aspect of the invention there are provided isolated polynucleotides (including genomic DNA, genomic RNA, cDNA and mRNA; double stranded as well as +ve and -ve strands) which encode the polypeptides of the invention. Single stranded DNA molecules of all or part of the ERG8 gene either +ve or -ve strand, find use inter alia, as
10 hybridisation probes or PCR amplification primers. The sense strand of the complete gene sequence of native ERG8 is depicted in Figure 1 (SEQ ID No. 5) hereinafter. It will be appreciated that a polynucleotide of the invention may comprise any of the degenerate codes for a particular amino acid, including the use of rare codons. Indeed, when producing the polypeptide by recombinant expression in heterologous host strains, it may be desirable to
15 adopt the codon usage (preference) of the host organism (Murray. N.A.R. 17:477-508, 1989).

Thus, according to a further aspect invention there is provided an isolated polynucleotide comprising nucleic acid encoding the amino acid sequence depicted in SEQ ID No. 7 or a variant thereof, such as one possessing at least 80% identity thereto.

The invention further comprises convenient fragments of any one of the above
20 polynucleotide/nucleic acid sequences. Convenient fragments may be defined by restriction endonuclease digests of nucleic acid comprising the ERG8 gene sequence. Such fragments are useful inter alia, for expressing short polypeptides fragments of ERG8 protein of the invention as well as for use as hybridisation probes. The present invention also provides a polynucleotide probe comprising any one of the above sequences or fragments together with a
25 convenient label or marker, preferably a non-radioactive label or marker. Following procedures well known in the art, the probes can be used to identify and isolate not only corresponding nucleic acid sequences (i.e *C. albicans* ERG8 gene sequences) but, if sufficiently homologous, can also be used to identify the analogous gene from other organisms using techniques well known to the person skilled in the art. Such sequences may be
30 comprised in libraries, such as genomic or cDNA libraries. The present invention also provides RNA transcripts corresponding to any of the above *C. albicans* ERG8 sequences or fragments. RNA transcripts can be used to prepare a polypeptide of the invention by *in vitro* translation techniques according to known methods (Sambrook *et al.* "Molecular Cloning- A

Laboratory Manual, second edition 1989"). The invention further comprises full-length or fragment lengths of ERG8 gene (coding sequence) flanked by non-coding sequence which may include natural or non-natural sequence containing restriction enzyme recognition sequence motifs. The incorporation of suitable restriction enzyme recognition sites either side of the ERG8 coding region, or indeed any polynucleotide sequence from ERG8, facilitates cloning of the ERG8 gene or polynucleotide sequence into a suitable vector. A suitable polynucleotide comprises a full length *C. albicans* ERG8 gene (encoding the polypeptide that starts with methionine at position 1 and terminates with the leucine that precedes the stop codon TAA at position 1299 of Figure 1) flanked by unique HindIII (5'-end)-XhoI (3'-end) restriction sites. Examples of oligonucleotide primers which are suitable for use in PCR amplification of ERG8, and which incorporate useful restriction enzyme sites to facilitate cloning, are disclosed as SEQ ID Nos. 10 and 11. Nucleotide changes or mutations may be introduced into a polynucleotide sequence by *de novo* polynucleotide synthesis, by site directed mutagenesis using appropriately designed oligonucleotide primers or by any other convenient means known to the person skilled in the art.

For expression purposes, it may be advantageous to engineer a restriction site at the 5'-end which is also capable of reconstituting the native amino-terminal methionine of the protein. The cleavage recognition sequence for the NcoI restriction enzyme not only includes a sequence that codes for methionine, but also one that is capable of retaining a functional Kozak consensus sequence, enabling the ERG8 gene to be cloned at the 3'-end of a suitable promoter element in an expression vector.

The polynucleotides can be synthesised chemically, or isolated by one of several approaches known to the person skilled in the art such as polymerase chain reaction (PCR) or ligase chain reaction (LCR) or by cloning from a genomic or cDNA library.

Once isolated or synthesised, a variety of expression vector/host systems may be used to express ERG8 coding sequences. These include, but are not limited to microorganisms such as bacteria expressed with plasmids, cosmids or bacteriophage; yeasts transformed with expression vectors; insect cell systems transfected with baculovirus expression systems; plant cell systems transfected with plant virus expression systems, such as cauliflower mosaic virus; or mammalian cell systems (for example those transfected with adenoviral vectors); selection of the most appropriate system is a matter of choice.

Expression vectors usually include an origin of replication, a promoter, a translation initiation site, optionally a signal peptide, a polyadenylation site, and a transcription

termination site. These vectors also usually contain one or more antibiotic resistance marker gene(s) for selection. As noted above, suitable expression vectors may be plasmids, cosmids or viruses such as phage or retroviruses. The coding sequence of the polypeptide is placed under the control of an appropriate promoter, control elements and transcription terminator so
5 that the nucleic acid sequence encoding the polypeptide is transcribed into RNA in the host cell transformed or transfected by the expression vector construct. The coding sequence may or may not contain a signal peptide or leader sequence for secretion of the polypeptide out of the host cell. Expression and purification of the polypeptides of the invention can be easily performed using methods well known in the art (for example as described in Sambrook et al.
10 "Molecular Cloning- A Laboratory Manual, second edition 1989").

The vectors containing the DNA coding for the ERG8 polypeptides of the invention can be introduced (i.e transformed or transfected) into *E. coli*, *S. cerevisiae*, *Pichia pastoris* or any other suitable host to facilitate their manipulation (i.e. for mutagenesis, cloning or expression). Performance of the invention is neither dependent on nor limited to any
15 particular strain of host cell or vector; those suitable for use in the invention will be apparent to, and a matter of choice for, the person skilled in the art.

Host cells transformed or transfected with a vector containing an ERG8 nucleotide sequence may be cultured under conditions suitable for the expression and recovery of the encoded proteins from the cell culture. Such expressed proteins/polypeptides may be secreted
20 into the culture medium or they may be contained intracellularly depending on the sequences used, i.e. whether or not suitable secretion signal sequences were present.

The full-length native isolated *C. albicans* ERG8 protein (PMK enzyme) of the present invention, or a functional variant thereof, is useful as a target in biochemical assays, particularly for use in identifying inhibitors of the enzyme. However, to provide sufficient
25 enzyme for a biochemical assays (for example, for use in a high throughput screen for enzyme inhibitors) the enzyme has to be expressed at high levels and it has to be purified. Two major constraints impair ERG8 expression and purification: (i) ERG8 is not expressed at high levels from *C. albicans*, and (ii) expression and protein purification methodology is not well advanced for *C. albicans*.

30 We have now been able to overcome these problems by controlled over-expression of the *C. albicans* ERG8 in a strain of *Saccharomyces cerevisiae*. *S. cerevisiae* is a model system for expression and purification of recombinant proteins. Use of *S. cerevisiae* to express *C. albicans* ERG8 means that transformation, expression and purification

methodology used to produce and isolate the ERG8 protein can follow published procedures. As stated above, the invention is not limited to use of *S. cerevisiae* as the host for expression of *C. albicans* ERG8.

According to a further aspect of the invention there is provided a host cell adapted to
5 express *C. albicans* ERG8 polypeptide or a variant thereof. The yeast *S. cerevisiae* is the preferred host cell of choice. According to a further aspect of the invention there is provided a novel expression system for expression of the *C. albicans* ERG8 gene, which system comprises an *S. cerevisiae* host strain having the *C. albicans* ERG8 gene in place of the native ERG8 gene from *S. cerevisiae*, whereby the *C. albicans* ERG8 gene is expressed. Preferred *S.*
10 *cerevisiae* strains include JK9-3Dα and its haploid segregants.

The *C. albicans* ERG8 gene is preferably over-expressed relative to the expression derived from its own promoter. This is conveniently achieved by replacing the *C. albicans* ERG8 promoter by a stronger and preferably inducible promoter such as the *S. cerevisiae* GAL1 promoter, alpha factor or alcohol oxidase (for reviews see Ausubel et al. "Current
15 Protocols in Molecular Biology", John Wiley & Sons, New York.).

The novel expression system is conveniently prepared by transformation of a heterozygous ERG8 deletion strain of a convenient *S. cerevisiae* host by a suitable plasmid comprising the *C. albicans* ERG8 gene using methods well known in the art (Ito et al. J. Bacteriol. 153:163-168,1983; Schiestl and Grietz, Current Genetics 16:339-346,1989).

20 The plasmid comprising the *C. albicans* ERG8 represents a further aspect of the invention. Particularly suitable plasmids for expression of *C. albicans* ERG8 in *S. cerevisiae* include pYES2(Invitrogen) and plasmids derived from pYES2 carrying a native *S. cerevisiae* promoter such as the glyceraldehyde-3-dehydrogenase promoter.

The heterozygous ERG8 deletion strain of a diploid *S. cerevisiae* host is conveniently
25 achieved by disruption preferably using an antibiotic resistance cassette such as the kanamycin resistance cassette described by Wach et al (Yeast. 10:1793-1808, 1994).

As described earlier, the *C. albicans* ERG8 enzyme may be used in biochemical assays to identify agents which modulate the activity of the enzyme. The design and implementation of such assays will be evident to the biochemist of ordinary skill. The enzyme may be used to
30 turn over a convenient substrate whilst incorporating/losing a labelled component to define a test system. Test compounds are introduced into the test system and measurements made to

determine their effect on enzyme activity. Such assays are useful to identify inhibitors of the enzyme which may then prove valuable as antifungal agents.

Thus, in a further aspect of the invention we provide the use of a *C. albicans* ERG-8 gene and/or *C. albicans* PMK enzyme in an assay to identify inhibitors of the enzyme. In particular, we provide their use in pharmaceutical or agrochemical research.

Thus, according to a further aspect of the invention there is provided a method of identifying compounds that modulate, preferably inhibit, the activity of phosphomevalonate kinase (PMK), comprising, contacting a test compound with a polypeptide of the invention and determining the effect that the test compound has on the activity of the polypeptide.

10 The PMK (ERG8) protein catalyses the conversion of phosphomevalonate + ATP to pyrophosphomevalonate + ADP. By way of non-limiting example, the activity of the ERG8 enzyme may be determined by (i) measuring the increase in ADP production, (ii) by following the loss of ATP, or (iii) by monitoring transfer of radioactive label (i.e H^3 , C^{14} , P^{32}) into phosphomevalonate.

15 A suitable assay that measures ADP production involves coupling the ADP produced by the action of PMK on phosphomevalonate + ATP substrate with pyruvate kinase and phosphoenolpyruvate to form pyruvate and ATP. The pyruvate is then reduced to lactate with lactate dehydrogenase which converts NADH to NAD. The production of NAD (directly linked to ADP production indicative of PMK action) is conveniently measured by detecting
20 the change in absorbance at 340nm (NADH oxidation product). In this assay, test compounds that inhibit PMK activity are identified by determining the ability of a compound to inhibit PMK activity as assessed by a reduction in ADP production as gauged by a reduction in the production of NAD from NADH using pyruvate kinase and lactate dehydrogenase as coupling enzymes as described above. The person skilled in the art would be able to develop other
25 assays for measuring PMK activity without inventive input.

ATP can be conveniently assayed using commercially available kits (i.e Boehringer Mannheim) to monitor luminescence resulting from oxidation of luciferin to luciferase (Ford et al. J. Biolumin. Chemilumin. 11:149-167, 1996).

A suitable reaction that measures the production of radioactively labelled
30 phosphomevalonate involves incubation PMK enzyme with cofactors, substrate ATP and phosphomevalonate, one of which carries a radioactive label. After reaction, pyrophosphomevalonate can be resolved from unreacted substrate by high voltage electrophoresis at pH3.5 on 3MM paper and the amount of radioactivity incorporated into

pyrophosphomevalonate can be measured by scintillation counting (Lee and O'Sullivan. J. Biol. Chem. 260:13909-13915, 1985).

Any convenient test compound or library of test compounds may be used in conjunction with the test assay. Particular test compounds include low molecular weight
5 chemical compounds (preferably with a molecular weight less than 1500 daltons) suitable as pharmaceutical or veterinary agents for human or animal use, or compounds for non-administered use such as cleaning/sterilising agents or for agricultural use.

The ERG8 enzyme of the invention, and convenient fragments thereof may be used to raise antibodies. Such antibodies have a number of uses which will be evident to the
10 molecular biologist or immunologist of ordinary skill. Such uses include, but are not limited to, monitoring enzyme expression, development of assays to measure enzyme activity, precipitation or purification of the enzyme and as a diagnostic tool to detect *C. albicans*. Enzyme linked immunosorbant assays (ELISAs) are well known in the art and would be particularly suitable for detecting the ERG8 polypeptide or fragments thereof. Antibodies
15 raised against the polypeptides of the invention may be polyclonal, obtained for example by injecting the polypeptide(s) into a selected mammal (i.e. rabbit, mouse, goat or horse), and later collecting the immunised serum from the animal, and treating this according to procedures known in the art. Depending on the host species, various adjuvants may be used to enhance the immunological response against the injected polypeptide. Suitable adjuvants
20 include, but are not limited to Freud's, aluminium hydroxide and SAF. Antibodies may also be monoclonal antibodies produced by hybridoma cells, phage display libraries or other methodology. Monoclonal antibodies may be inter alia, human, rat or mouse derived. For the production of human monoclonal antibodies, hybridoma cells may be prepared by fusing spleen cells from an immunised animal, e.g. a mouse, with a tumour cell. Appropriately
25 secreting hybridoma cells may thereafter be selected (Koehler & Milstein. Nature. 256:495-497, 1975; Cole et al. "Monoclonal antibodies and Cancer Therapy, Alan R Liss Inc, New York N.Y. pp 77-96). Rodent antibodies may be humanised using recombinant DNA technology according to techniques known in the art. Alternatively, chimeric antibodies, single chain antibodies, Fab fragments may also be developed against the polypeptides of the
30 invention (Huse et al. Science. 256:1275-1281, 1989), using skills known in the art.

The polynucleotides and antibodies of the invention may be used in gene-probe or protein-probe methodologies, with or without amplification (for example, via PCR or second antibody detection) to detect or diagnose the presence of *C. albicans*. This is particularly

valuable in diagnosing clinical infections. Accordingly, the invention provides diagnostic kits for the detection of *C. albicans* ERG8 or fragments thereof, and provides for the use of ERG8 protein, polypeptide fragments thereof and/or antibodies raised thereagainst as positive control. The reagents in the kit may be compartmentalised and the kit may also comprise
5 instructions for use.

DNA diagnostics is based on DNA/RNA hybridisation technology, i.e. the specific *in vitro* binding of complementary single-stranded nucleic acid with the formation of double-stranded nucleic acid. The DNA/DNA or DNA/RNA double strands formed are termed hybrids. To detect the presence of *C. albicans* in a bodily fluid such as blood, total nucleic
10 acid is isolated from the test fluid sample using standard techniques and the presence of *C. albicans* ERG8 nucleic acid in the sample is detected using for example detectably labelled probes comprising one or more of the polynucleotides of the invention. The probes can be short, chemically synthesised oligonucleotide probes of a length of approximately 10 - 50 nucleotides, or may be recombinantly expressed fragments of the ERG8 gene of
15 approximately 0.3-1.5Kb in size. Single stranded oligonucleotide probes which are specific for *C. albicans* are preferred. The probe can be provided with a suitable detectable reporter molecule label such as a radioisotope (P^{32} , tritium, C^{14} or S^{35}), or a non-radioactive label such as digoxigenin or biotin, using techniques available to the person skilled in the art. Prior to the hybridisation reaction, all or any part of *C. albicans* ERG8 DNA containing the sequence to
20 which the probe can hybridise, present in the test sample is amplified using for example PCR (polymerase chain reaction) or LCR (ligase chain reaction). For the specific hybridisation reaction, the test nucleic acid and if necessary the probe DNA is converted into single strands by denaturation (heat or alkali) and then very specifically hybridised with each other under stringent conditions. Under appropriate conditions the gene probe only hybridises to
25 complementary sequences of the DNA or RNA to be detected. The hybridisation and detection assay can be carried out in a number of different formats known to the person skilled in the art including, solid-phase hybridisation of target DNA or probe coupled to a solid support such as nitrocellulose or magnetic beads. The hybridisation complex can then be determined quantitatively, following removal of unbound probe or test nucleic acid, by way of
30 the reporter molecule label (e.g. fluorescent or radioactive) employed.

The test sensitivity of this single gene-probe diagnostic method can be increased by combination with DNA or RNA amplification techniques such as PCR or LCR. Using such amplification techniques, the DNA to be detected can be multiplied by up to 10^9 .

There may only be 100-1000 organisms per ml of blood in association with *Candida* infections. Such small numbers of cells are easily detectable when combining the amplification and DNA-probe detection techniques offering the possibility of early detection of infection.

5 Thus, according to a further aspect of the invention there is provided a method of diagnosing the presence of the *C. albicans* ERG8 gene in a test sample, comprising: contacting a polynucleotide probe of at least 15 nucleotides in length, which probe is capable of specifically hybridising with the sequence depicted in SEQ ID No. 6, with the test sample under conditions which allow duplex formation between said polynucleotide probe and the
10 nucleic acid in the test sample; and, detecting duplex formation. In a preferred embodiment the polynucleotide probe is detectably labelled. In another embodiment the polynucleotide probe is single stranded. In another embodiment the polynucleotide probe is completely complementary to the target sequence to be detected. According to a further aspect of the invention the polynucleotide probe is substituted for by a pair of oligonucleotide primers
15 capable of specific PCR amplification of all or part of the ERG8 gene in the test sample, with subsequent identification of amplification product.

According to another aspect of the present invention there is provided a diagnostic kit for diagnosing or detecting the presence of *C. albicans* comprising, one or more diagnostic probe(s) and/or diagnostic primer(s) and/or antibodies capable of selectively hybridising or
20 binding to the polynucleotide of SEQ ID No. 6 or the polypeptide of SEQ ID No. 7, or to variant sequences thereof as defined herein.

In a preferred embodiment, the diagnostic (detection) probes are provided on a microarray.

Such kits may further comprise appropriate buffer(s) and/or polymerase(s) such as
25 thermostable polymerases, for example taq polymerase. They may also comprise companion/constant primers and/or control primers or probes. A companion/constant primer is one that is part of the pair of primers used to perform PCR. Such primer usually complements the template strand precisely.

In another embodiment the kit is an ELISA kit comprising one or more antibodies
30 specific for the polypeptide depicted in SEQ ID No. 7, or a variant thereof as defined herein.

The following examples and figure describe and illustrate the invention. They are not intended to limit the scope of the invention in any way:

Figure 1 shows the nucleotide sequence of the *C. albicans* gene encoding phosphomevalonate kinase. Translation start (ATG) and stop (TAA) codons are highlighted.

Examples

5

1. Cloning and partial sequence determination of two separate clones from a *Candida albicans* genomic library.

Two separate cloned and sequenced nucleic acid sequences from a *C. albicans* library (SEQ ID NOs. 1 & 3) were found to have homology to that of *S. cerevisiae* ERG8 gene. The
10 complement of specific regions in SEQ ID Nos. 1 and 3 were synthesised as oligonucleotides (SEQ ID Nos. 2 & 4) for use in the isolation of a clone containing the *C. albicans* ERG8 gene.

2. Cloning and sequence determination of *Candida albicans* ERG8.

Using the two oligonucleotide primers (SEQ ID Nos. 2 and 4), the *C. albicans* ERG8
15 gene was isolated as a plasmid clone from a library of *C. albicans* genomic DNA in the yeast shuttle vector YEp24 using PCR. The *C. albicans* library was maintained in *E. coli* and independent bacterial colonies were grown in single wells of each of 15 x 384-well microtitre plates. The properties of the library plasmids are such that this gridded array contains approximately 2.5x the amount of DNA in the *C. albicans* genome.

20 Small aliquots of cells from each of the wells were mixed to produce a pool of cells that were derived from all of the wells from a single plate. Similar pools were made for all of the rows and all of the columns from each of the plates. Samples of each of the pools of the cells for each complete plate were used in PCR reactions with SEQ ID Nos. 2 and 4 oligonucleotide primers to identify plate(s) in the array carrying *C. albicans* ERG8.

25 Subsequent PCR reactions with pools of cells from rows of wells and columns of wells defined the specific well(s) carrying a clone of *C. albicans* ERG8.

The PCR reactions contained in a total volume of 0.05ml: 75mM Tris-HCl (pH 8.8 at 25°C), 20mM (NH₄)₂SO₄, 1.5mM MgCl₂, 0.01% Tween 20, 0.2mM of each of dATP, dCTP, dGTP and dTTP, 1.25 units Taq DNA polymerase, 100pmoles of each oligonucleotide primer
30 and 0.005ml *E. coli* cell suspension. PCR reactions were incubated at 94°C for 1 min then for 30 cycles of the following: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min. PCR products were analysed by electrophoresis through agarose and visualised under UV light after staining with ethidium bromide.

Putative clones harbouring the ERG8 gene were selected, the plasmid DNAs in these clones were purified and the complete sequence of the *C. albicans* ERG8 gene was determined on both strands using flanking sequence- or insert sequence-specific oligonucleotide primers. The full-length of the *C. albicans* ERG8 gene, including 'start' ATG and 'stop' TAA is shown in Figure 1. The protein translation of the gene is depicted in SEQ ID No. 7.

3. Generation of a heterozygous ERG8 deletion strain of *S. cerevisiae*

Since PMK is an essential enzyme, only one allele of a diploid cell can be deleted without loss of viability. One ERG8 gene diploid strain of *S. cerevisiae* (JK9-3daa; Kunz et al., Cell 73:585-596 (1993)) was disrupted using a kanamycin resistance cassette as described by Wach et al. (Yeast 10:1793-1808, 1994) using the protocol described therein with the oligonucleotides shown in SEQ ID Nos. 8 and 9. Sporulation of the heterozygous diploid (ERG8/erg8::KanMX) yields only two viable spores that are both sensitive to kanamycin, showing ERG8 to be essential, and the characteristic arrest phenotype for the two inviable spores.

4. Complementation of a *S. cerevisiae* ERG8 deletion with the cloned *C. albicans* ERG8

The heterozygous ERG8/erg8::KanMX strain was transformed with the plasmid carrying the full-length *C. albicans* ERG8 gene within a fragment of *C. albicans* genomic DNA such that expression of the gene will depend on functionality of the *C. albicans* promoter in the heterologous *S. cerevisiae* host. Surprisingly, the gene carried on the plasmid failed to complement the gene deletion as demonstrated by a failure to recover kanamycin-resistant haploid cells after sporulation. This was probably due to inappropriate expression of *C. albicans* ERG8 in *S. cerevisiae*.

To enable expression of *C. albicans* ERG8 in *S. cerevisiae* and to facilitate purification of ERG8 protein as a result of over-expression in a suitable host, the *C. albicans* promoter was replaced by the efficient, inducible *S. cerevisiae* GAL1 promoter. The *C. albicans* ERG8 coding sequence was amplified by PCR using the oligonucleotides shown in SEQ ID Nos. 10 and 11, which contain convenient restriction enzyme sites for cloning the product of PCR into an appropriate expression vector such as pYES2 (Invitrogen). The identity of the PCR-amplified gene cloned into pYES2 was confirmed by DNA sequencing. After transformation into the heterozygous ERG8/erg8::KanMX strain, the plasmid was able

to complement the *erg8::KanMX* allele in *S. cerevisiae* since kanamycin-resistant haploid spores were viable on medium containing galactose but not glucose. This *S. cerevisiae* strain is a useful source of biologically active *C. albicans* ERG8 protein for assays *in vitro*.

C. albicans ERG8 can also be conveniently over-expressed in bacteria such as *E. coli*.

5 The *C. albicans* ERG8 coding sequence is amplified by PCR using oligonucleotides containing convenient restriction sites for cloning into expression vectors such as pT7#3.3. It is particularly convenient if the initiation codon for ERG8 is incorporated within one of the restriction sites. Oligonucleotides suitable for this are shown in SEQ ID Nos. 12 and 13. Oligonucleotides may also incorporate extra sequences to encode a small "tag" that aids the
10 subsequent purification of the protein. Such tags include for example, the "His₆" tags which may be incorporated at the N- or C-terminus of ERG8 using the oligonucleotides shown in SEQ ID Nos. 14 and 15. Recombinantly expressed tagged ERG8 protein can be conveniently purified by affinity chromatography purification methodology using commercially available purification kits (i.e Qiagen) (Borsig et al., Biochem. Biophys. Res. Commun. 240:586-589,
15 1997).

5. Hybridisation test of nucleic acid variations of specific nucleic acid sequences

5.1 Hybridisation Test

A method for detecting variant nucleic acids containing sequences related to specific
20 ERG8 sequences such as natural alleles, is described. These variant nucleic acids may be present in a variety of forms such as within plasmids or other like vehicles which may be fixed on to a hybridisation membrane, such as a nitrocellulose or nylon filter ready for detection using a labelled probe. Hybridisation assays can also be performed to identify variant sequences from within genomic or cDNA libraries. Hybridisation technology is well
25 advanced. It will be apparent to the person skilled in the art that the protocol described below is only one example of a hybridisation protocol suitable to identify ERG8 variant sequences.

5.2 Hybridisation probe

Hybridisation probes may be generated from any fragment of DNA or RNA encoding the specific ERG8 nucleic sequence of interest. Such fragments can be for example, restriction
30 fragments isolated following restriction enzyme digestion of nucleic acid containing the ERG8 nucleotide sequence or synthetic oligonucleotides specific for a region of the ERG8 gene or a complementary sequence thereto.

A hybridisation probe can be generated from a synthetic oligonucleotide or a dephosphorylated restriction fragment sequence by addition of a radioactive 5' phosphate group from $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ by the action of T4 polynucleotide kinase. 20 pmoles of the oligonucleotide are added to a 20 μl reaction containing 100mM Tris, pH7.5, 10mM MgCl_2 , 5 0.1mM spermidine, 20mM dithiothreitol (DTT), 7.55 μM ATP, 55 μCi $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and 2.5u T4 polynucleotide kinase (Pharmacia Biotechnology Ltd, Uppsala, Sweden). The reaction is incubated for 30 minutes at 37°C and then for 10 minutes at 70°C prior to use in hybridisation. Methods for the generation of hybridisation probes from oligonucleotides or from DNA and RNA fragments (Chapters 11 and 10 respectively in Sambrook et al. *ibid*). A 10 number of proprietary kits are also available for these procedures.

5.3 Hybridisation conditions

Filters containing the nucleic acid are pre-hybridised in 100ml of a solution containing 6x SSC, 0.1%SDS and 0.25% dried skimmed milk (Marvel™) at 65°C for a minimum of 1 hour in a suitable enclosed vessel. A proprietary hybridisation apparatus such 15 as model HB-1 (Techne Ltd) provides reproducible conditions for the experiment.

The pre-hybridisation solution is then replaced by 10ml of a probe solution containing 6xSSC, 0.1% SDS, 0.25% dried skimmed milk (e.g. Marvel™) and the oligonucleotide probe generated above. The filters are incubated in this solution for 5 minutes at 65°C before allowing the temperature to fall gradually to below 30°C. The probe solution 20 is then discarded and the filters washed in 100ml 6xSSC, 0.1% SDS at room temperature for 5 minutes. Further washes are then made in fresh batches of the same solution at 30°C and then in 10°C increments up to 60°C for 5 minutes per wash.

After washing, the filters are dried and used to expose an X-ray film such as Hyperfilm™ MP (Amersham International) at -70°C in a light-tight film cassette using a fast 25 tungstate intensifying screen to enhance the photographic image. The film is exposed for a suitable period (normally overnight) before developing to reveal the photographic image of the radio-active areas on the filters. Related nucleic acid sequences are identified by the presence of a photographic image compared to totally unrelated sequences which should not produce an image. Generally, related sequences will appear positive at the highest wash

temperature (60°C). However, related sequences may only show positive at the lower wash temperatures (50, 40 or 30°C).

These results will also depend upon the nature of the probe used. Longer nucleic acid fragment probes will need to be hybridised for longer periods at high temperature but
5 may remain bound to related sequences at higher wash temperatures and/or at lower salt concentrations. Shorter, mixed or degenerate oligonucleotide probes may require less stringent washing conditions such as lower temperatures and/or higher Na⁺ concentrations. A discussion of the considerations for hybridisation protocols is provided in Sambrook et al. (Chapter 11).

10 To prepare 20 x SSC, 175.3 g of NaCl and 88.2 g of sodium citrate is dissolved in approximately 800ml of water, the pH is adjusted to 7.0 using 10 N solution of NaOH and the volume is adjusted to 1 litre with water, before autoclaving.

- 20 -

Claims:

1. A purified polypeptide comprising the amino acid sequence depicted in SEQ ID No. 7 or a sequence possessing at least 80% similarity thereto.
2. An isolated polypeptide of at least 15 contiguous amino acids of the polypeptide of claim 1.
- 5 3. An antibody specific for the polypeptide of claim 1 or 2.
4. An antibody as claimed in claim 3 which is a monoclonal antibody.
5. A purified polynucleotide comprising a nucleic acid sequence encoding the polypeptide depicted in SEQ ID No. 7 or a sequence possessing at least 80% identity thereto.
6. A polynucleotide of at least 15 nucleotides in length, which polynucleotide is capable of
10 specifically hybridising to a nucleic acid sequence selected from the group consisting of SEQ ID Nos. 1, 3, 5 or 6, or a sequence complementary to any of said sequences.
7. An expression vector comprising the polynucleotide of claim 5.
8. A host cell which contains an expression vector according to claim 7.
9. A method for producing the polypeptide of claim 1, comprising:
15 (a) culturing a host cell according to claim 8 under conditions suitable for the expression of said polypeptide, and
(b) recovering said polypeptide from the host cell or cell culture.
10. Use of the polypeptide of claim 1 in an assay to identify compounds that inhibit phosphomevalonate kinase (PMK) activity.
- 20 11. A method of identifying compounds that modulate the activity of PMK, comprising:
(a) contacting a test compound with a polypeptide according to claim 1, and
(b) determining the effect that the test compound has on the activity of the polypeptide.
12. A compound identified by the method of claim 11.
13. A method for detecting or diagnosing the presence of *Candida albicans* in a test sample,
25 comprising contacting the sample with an agent capable of detecting a polypeptide possessing the amino acid sequence depicted in SEQ ID No. 7 or a sequence possessing at least 80% similarity thereto, or a nucleic acid sequence encoding the polypeptide depicted in SEQ ID No. 7 or a sequence possessing at least 80% identity thereto.
14. A method as claimed in claim 13 wherein the presence of the nucleic acid is detected
30 using an oligonucleotide primer or probe capable of selectively hybridising to the said polynucleotide.
15. A diagnostic kit for detecting the presence of *C. albicans* comprising: one or more diagnostic probe(s) and/or diagnostic primer(s) and/or antibodies capable of selectively

- 21 -

hybridising or binding to the polynucleotide of claim 6 or the polypeptide of claim 1, and instructions for use.

1 / 1

GTGGAAAAAAGACAGAACAGTAGATTCCAACCTTCAGAAATATTCATTGATCTGAACATTT
CTTTTTCTCCGATCATCAATTGGCAATGTCAAAGCATTAGTGCACCTGGAAAAGCATTCT
TGCTGGTGGATATTTGGTTCTTGAGCCAATTTATGATGCTTATGTGACAGCATTGTCATCAG
AATGCATGCAGTTATAACACCAAAAGGAACCAGTTTGAAAGAATCTAGAATCAAATTTCTTC
ACCCCAATTTGCAAACGGAGAATGGGAATATCACATATCATCAAATACAGAGAAGCCCAGAGA
AGTTCAGTCACGCATAAATCCATTTTTAGAGGCAACTATATTCATCGTTTTAGCTTATATTCA
ACCGACCGAAGCATTGATCTTGAAATCATCATTTACTCAGACCCTGGATATCATTCAAGA
AGATACTGAAACCAAGACATCCTCGAATGGAGAAAAACATTTCTTTACCATTCTCGTGCCAT
TACCGAAGTGGAAAAGACCGGATTAGGTTTCATCGGCAGGATTAGTGTGAGTTGTTGCCACAAG
TTTATTATCCCATTTTTATCCCAATGTTATCAGTACGAATAAAGATATTTGACACAACGTTGC
ACAGATTGCACATTGTTATGCCCAAAAAAGATAGGATCTGGGTTTGATGTTGCAACTGCAAT
TTATGGTCTGATTGTATATAGAAGATTTAGCCAGCTTTGATAAATGACGTGTTTCAGGTCT
AGAAAGTGATCCTGAGAAGTTCCCCACAGAGTTGAAAAAATTGATTGAAAGTAACTGGGAATT
CAAACATGAAAGATGTACATTACCATACGGAATCAAGTTATTAATGGGTGACGTCAAGGGTGG
CTCAGAAACACCCAAATTGGTATCACGAGTACTCCAATGGAAAAAGGAAAAGCCAGAAGAAAG
CTCTGTTGTGTATGACCAGCTTAATAGTGCCAATTTACAGTTTATGAAGGAATTGAGGGAAAT
GCGTGAAAAATACGACTCAGACCCAGAGACTTATATTAAAGAGTTAGATCATTCTGTTGAGCC
TTTGAAGTTGCGGATTAAGAACATCAGAAAAGGGTTACAAGCATTAAACACAAAAATCAGAGGT
TCCAATTGAACCTGATGTCCAAACCCAGTTGTTGGACCGTTGTCAAGAGATTCCTGGTTGTGT
TGGTGGTGTGGTTCCAGGTGCTGGTGGATACGATGCAATAGCTGTATTAGTGTGGAAAATCA
AGTGGGAAATTTTAAGCAGAAAACCTTTGAAAATCCAGATTATTTTCATAATGTTTACTGGGT
TGATTTGGAAGAGCAAACAGAAGGTGTACTTGAAGAAAAACCAGAAGACTATATAGGTTTATA
AAATATCACTGGGATATGTCTACAAGGTGTTTTCGATTAGAGTTTTTGATCCCCATTTTAACA
TATTTTACTTCAATCTTACACTTTATCCTTTTAAGTAGGTATGTGTAGGGAAAGAGCCTGATC
TTCATAAACCGTTGCAAACTAATTGATTATATTTTCTATTGTAAATTTTCATATGCAGGAAATA
GCTTATTCGACAAATTATTTATTTTCGTCTCGTTCTGGTCCAAGTACCCAGAGACGAAATAA
CTGACAACACGCAGGGCTGGGTGGCATTTTCGTACACGATTATTATTAATGGTAACAAAAA
AAGGGGRKATGCCCGTGGTCGATACACAAATATTTATGATATACTTTCCATATTTTTTTTTT

FIGURE 1

- 1 -

SEQUENCE LISTING

<110> ASTRAZENECA AB

5 <120> PROTEIN

<130> LDSG/PHM70579/WO

<140>

10 <141>

<150> GB 9919766.7

<151> 1999-08-21

15 <160> 11

<170> PatentIn Ver. 2.1

<210> 1

20 <211> 547

<212> DNA

<213> Candida albicans

<400> 1

25 ccaatggaaa aaggaaaagc cagaagaaag ctctgttggtg tatgaccagc ttaatagtgc 60
 caattttacag tttatgaagg aattgagggg aatgcgtgaa aaatacgact cagacccaga 120
 gacttatatt aaagagtttag atcattctgt tgagcctttg actggttgcga ttaagaacat 180
 cagaaaaagg ttacaagcat taacacaaaa atcagagggtt ccaattgaac ctgatgtcca 240
 aaccagttg ttggaccgtt gtcaagagat tcctgggtgt gttggcgtg tggttccagg 300
 30 tgctggtgga tacgatgcaa tagctgtatt agtgttgga aatcaagtgg gaaatttta 360
 gcagaaaact cttgaaaatc cagattatct tcataatgtt tactgggttg atttggaaga 420
 gcaaacagaa ggtgtacttg aagaaaaacc agaagactat ataggtttat aaaatatcac 480
 taggatatgt ctacaagggtg atttcgatta gattttctgc taccggtttt aacatatttt 540
 acttcaa 547

35

<210> 2

<211> 21

<212> DNA

40 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Single-stranded oligonucleotide

45 <400> 2

gctggtggat acgatgcaat a 21

- 2 -

<210> 3
 <211> 577
 <212> DNA
 <213> Candida albicans

5

<400> 3
 atgtacatct ttcattgttg aattcccagt tacttgcaat caattttttc aactctgtgg 60
 ggaacttctc aggatcactt tctagaacct gaaacacgtc atttatcaaa gctggctgaa 120
 atcttctata tacaatcaga ccataaattg cagttgcaac atcaaaccga gatcctatct 180
 10 ttttttgggc ataacaatgt gcaatctgtg caacgttgtg caaaatatct ttattcgtac 240
 tgataacatt ggggataaaa tgggataata aacttggtgc aacaactgac actaatcctg 300
 ccgatgaacc taatccggtc ttttccactt cggtaatggc acgagaatgg taaagaaaag 360
 ttttttctcc attcaggat gtcttggttt cagtatcttc ttgtgaatga tatccagggt 420
 ccgagtaaat aatgatttca agatcaaatg cttcggtcgg ttgaatataa gctaaaaacc 480
 15 gatggatata gttgcctcta aaaatgggat ttatgcgtga ctgnacttct ttgggttttc 540
 ngtaattgat gatattgat antccattc cccggtt 577

<210> 4
 20 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 25 <223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 4
 ggggataaaa tgggataata aactt 25

30

<210> 5
 <211> 1763
 <212> DNA
 <213> Candida albicans

35

<400> 5
 gtggaaaaaa aagacagaac agtagattcc aacttcagaa tattcattca gatctgaaca 60
 tttctttttc tccgatcatc aattggcaat gtcaaaagca tttagtgcac ctggaaaagc 120
 atttcttgct ggtggatatt tggttcttga gccaatattat gatgcttatg tgacagcatt 180
 40 gtcacacga atgcattgag ttataacacc aaaaggaacc agtttgaaag aatctagaat 240
 caaaatttct tcaccccaat ttgcaaacgg agaattggaa tatcacatat catcaaatac 300
 agagaagccc agagaagtgc agtcacgcat aaatccattt ttagaggcaa ctatattcat 360
 cgttttagct tatattcaac cgaccgaagc atttgatctt gaaatcatca tttactcaga 420
 ccctggatat cattcacaag aagatactga aaccaagaca tcctcgaatg gagaaaaaac 480
 45 atttctttac cattctcgtg ccattaccga agtggaaaag accggattag gttcatcggc 540
 aggattagtg tcagttgttg ccacaagttt attatcccat tttatcccca atgttatcag 600
 tacgaataaa gatattttgc acaacgttgc acagattgca cattgttatg cccaaaaaaa 660
 gataggatct ggggtttgat ttgcaactgc aatttatggg ctgattgtat atagaagatt 720

- 3 -

tcagccagct ttgataaatg acgtgtttca ggttctagaa agtgatcctg agaagttccc 780
 cacagagttg aaaaaattga ttgaaagtaa ctgggaattc aaacatgaaa gatgtacatt 840
 accatacggg atcaagttat taatgggtga cgtcaagggt ggctcagaaa caccctaat 900
 ggtatcacga gtactccaat ggaaaaagga aaagccagaa gaaagctctg ttgtgtatga 960
 5 ccagcttaat agtgccaatt tacagtttat gaaggaattg agggaaatgc gtgaaaaata 1020
 cgactcagac ccagagactt atattaaaga gttagatcat tctgttgagc ctttgactgt 1080
 tgcgattaag aacatcagaa aagggttaca agcattaaca caaaaatcag aggttccaat 1140
 tgaacctgat gtccaaaccc agttgttgga ccgttgtcaa gagattcctg gttgtgttg 1200
 tggtgtggtt ccaggtgctg gtggatacga tgcaatagct gtattagtgt tggaaaatca 1260
 10 agtgggaaat tttaagcaga aaactcttga aaatccagat tattttcata atgtttactg 1320
 ggttgatttg gaagagcaaa cagaaggtgt acttgaagaa aaaccagaag actatatagg 1380
 ttataaaaat atcactggga tatgtctaca aggtgttttc gattagagtt tttgatcccc 1440
 attttaacat attttacttc aatcttacac tttatccttt taagtaggta tgtgtaggga 1500
 aagagcctga ttttcataaa ccgttgcaaa ctaattgatt atattttcta ttgtaaattt 1560
 15 catatgcagg aaatagctta ttcgacaaat tatttatttt cgtctcgttc tggccaagt 1620
 accccagaga cgaaataact gacaacacgc agggctgggt tggcattttc gtcacacgat 1680
 tattattaat ggtaacaaaa aaaggggrka tgcccgtggt cgatacacia atatttatga 1740
 tatactttcc atattttttt ttt 1763

20

<210> 6

<211> 1299

<212> DNA

<213> Candida albicans

25

<400> 6

atgtcaaaaag catttagtgc acctggaaaa gcatttcttg ctggtggata tttggttctt 60
 gagccaattt atgatgctta tgtgacagca ttgtcatcac gaatgcatgc agttataaca 120
 ccaaaaggaa ccagtttgaa agaacttaga atcaaaattt cttcacccca atttgcaaac 180
 30 ggagaatggg aatatcacat atcatcaa atcagagaagc ccagagaagt tcagtcacgc 240
 ataaatccat ttttagaggc aactatattc atcgttttag cttatattca accgaccgaa 300
 gcatttgatc ttgaaatcat catttactca gacctggat atcattcaca agaagatact 360
 gaaaccaaga catcctcgaa tggagaaaaa acatttcttt accattctcg tgccattacc 420
 gaagtggaaa agaccggatt aggttcacgc gcaggattag tgtcagttgt tgccacaagt 480
 35 ttattatccc attttatccc caatgttatc agtacgaata aagatatttt gcacaacgtt 540
 gcacagattg cacattgtta tgcccaaaaa aagataggat ctgggtttga tgttgcaact 600
 gcaatttatg gctgattgt atatagaaga tttcagccag ctttgataaa tgacgtgttt 660
 caggttctag aaagtgatcc tgagaagttc cccacagagt tgaaaaaatt gattgaaagt 720
 aactgggaat tcaaacatga aagatgtaca ttaccatacg gaatcaagtt attaatgggt 780
 40 gacgtcaagg gtggctcaga aacacccaaa ttggtatcac gagtactcca atggaaaaag 840
 gaaaagccag aagaaagctc tgttgtgtat gaccagctta atagtgccaa tttacagttt 900
 atgaaggaaat tgagggaat gcgtgaaaaa tacgactcag acccagagac ttatattaaa 960
 gagttagatc attctgttga gcctttgact gttgcgatta agaacatcag aaaagggtta 1020
 caagcattaa cacaaaaatc agaggttcca attgaacctg atgtccaaac ccagttgttg 1080
 45 gaccgttgtc aagagattcc tgggtgtgtt ggtggtgttg ttccagggtc tgggtggatac 1140
 gatgcaatag ctgtattagt gttggaaaaa caagtgggaa attttaagca gaaaactctt 1200
 gaaaatccag attattttca taatgtttac tgggttgatt tggaagagca aacagaaggt 1260
 gtacttgaag aaaaaccaga agactatata ggtttataa 1299

- 4 -

<210> 7
 <211> 432
 5 <212> PRT
 <213> Candida albicans

<400> 7
 Met Ser Lys Ala Phe Ser Ala Pro Gly Lys Ala Phe Leu Ala Gly Gly
 10 1 5 10 15
 Tyr Leu Val Leu Glu Pro Ile Tyr Asp Ala Tyr Val Thr Ala Leu Ser
 20 25 30
 15 Ser Arg Met His Ala Val Ile Thr Pro Lys Gly Thr Ser Leu Lys Glu
 35 40 45
 Ser Arg Ile Lys Ile Ser Ser Pro Gln Phe Ala Asn Gly Glu Trp Glu
 50 55 60
 20 Tyr His Ile Ser Ser Asn Thr Glu Lys Pro Arg Glu Val Gln Ser Arg
 65 70 75 80
 Ile Asn Pro Phe Leu Glu Ala Thr Ile Phe Ile Val Leu Ala Tyr Ile
 25 85 90 95
 Gln Pro Thr Glu Ala Phe Asp Leu Glu Ile Ile Ile Tyr Ser Asp Pro
 100 105 110
 30 Gly Tyr His Ser Gln Glu Asp Thr Glu Thr Lys Thr Ser Ser Asn Gly
 115 120 125
 Glu Lys Thr Phe Leu Tyr His Ser Arg Ala Ile Thr Glu Val Glu Lys
 130 135 140
 35 Thr Gly Leu Gly Ser Ser Ala Gly Leu Val Ser Val Val Ala Thr Ser
 145 150 155 160
 Leu Leu Ser His Phe Ile Pro Asn Val Ile Ser Thr Asn Lys Asp Ile
 40 165 170 175
 Leu His Asn Val Ala Gln Ile Ala His Cys Tyr Ala Gln Lys Lys Ile
 180 185 190
 45 Gly Ser Gly Phe Asp Val Ala Thr Ala Ile Tyr Gly Leu Ile Val Tyr
 195 200 205
 Arg Arg Phe Gln Pro Ala Leu Ile Asn Asp Val Phe Gln Val Leu Glu

- 5 -

	210		215		220
	Ser Asp Pro Glu Lys Phe Pro Thr Glu Leu Lys Lys Leu Ile Glu Ser				
	225		230		235 240
5	Asn Trp Glu Glu Lys His Glu Arg Cys Thr Leu Pro Tyr Gly Ile Lys				
		245		250	255
	Leu Leu Met Gly Asp Val Lys Gly Gly Ser Glu Thr Pro Lys Leu Val				
10		260		265	270
	Ser Arg Val Leu Gln Trp Lys Lys Glu Lys Pro Glu Glu Ser Ser Val				
		275		280	285
15	Val Tyr Asp Gln Leu Asn Ser Ala Asn Leu Gln Phe Met Lys Glu Leu				
		290		295	300
	Arg Glu Met Arg Glu Lys Tyr Asp Ser Asp Pro Glu Thr Tyr Ile Lys				
	305		310		315 320
20	Glu Leu Asp His Ser Val Glu Pro Leu Thr Val Ala Ile Lys Asn Ile				
		325		330	335
	Arg Lys Gly Leu Gln Ala Leu Thr Gln Lys Ser Glu Val Pro Ile Glu				
25		340		345	350
	Pro Asp Val Gln Thr Gln Leu Leu Asp Arg Cys Gln Glu Ile Pro Gly				
		355		360	365
30	Cys Val Gly Gly Val Val Pro Gly Ala Gly Gly Tyr Asp Ala Ile Ala				
		370		375	380
	Val Leu Val Leu Glu Asn Gln Val Gly Asn Phe Lys Gln Lys Thr Leu				
	385		390		395 400
35	Glu Asn Pro Asp Tyr Phe His Asn Val Tyr Trp Val Asp Leu Glu Glu				
		405		410	415
	Gln Thr Glu Gly Val Leu Glu Glu Lys Pro Glu Asp Tyr Ile Gly Leu				
40		420		425	430
	<210> 8				
45	<211> 70				
	<212> DNA				
	<213> Artificial Sequence				

- 6 -

<220>

<223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 8

5 aaatgtcaga gttgagagcc ttcagtgcc cagggaaagc gttactagct gcagctgaag 60
cttcgtacgc 70

<210> 9

10 <211> 73

<212> DNA

<213> Artificial Sequence

<220>

15 <223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 9

20 agttatttat caagataagt ttccggatct ttttctttcc taacacccca ggcataggcc 60
actagtggat ctg 73

<210> 10

<211> 33

<212> DNA

25 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Single-stranded oligonucleotide

30 <400> 10

cccaagcttg gcaatgtcaa aagcatttag tgc 33

<210> 11

35 <211> 36

<212> DNA

<213> Artificial Sequence

<220>

40 <223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 11

ccgctcgaga ttttataaac ctatatagtc ttctgg 36

45

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14533 A3

(51) International Patent Classification⁷: C12N 15/54,
9/12, C12Q 1/68, G01N 33/50

John, David, Charles [GB/US]; 35 Gatehouse Drive,
Waltham, MA 02451 (US). SCHNELL, Norbert, Friede-
mann [DE/US]; 35 Gatehouse Drive, Waltham, MA 02451
(US).

(21) International Application Number: PCT/GB00/03100

(22) International Filing Date: 15 August 2000 (15.08.2000)

(74) Agent: GAINNEY, Laurence, David, Scott; Astrazeneca,
Global Intellectual Property, PO Box 272, Mereside, Alder-
ley Park, Macclesfield, Cheshire SK10 4GR (GB).

(25) Filing Language: English

(26) Publication Language: English

(81) Designated States (*national*): JP, MG, US.

(30) Priority Data:
9919766.7 21 August 1999 (21.08.1999) GB

(84) Designated States (*regional*): European patent (AT, BE,
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,
NL, PT, SE).

(71) Applicant (*for all designated States except MG, US*): AS-
TRAZENECA AB [SE/SE]; S-151 85 Sodertalje (SE).

Published:
— with international search report

(71) Applicant (*for MG only*): ASTRAZENECA UK LIM-
ITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN
(GB).

(88) Date of publication of the international search report:
27 September 2001

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): ROSAMOND,

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*



WO 01/14533 A3

(54) Title: PHOSPHOMEVALONATE KINASE (PMK) GENE (ERG8) FROM *CANDIDA ALBICANS*

(57) Abstract: This invention relates to polynucleotides, polypeptides encoded by these polynucleotides, to the production of such polynucleotides and polypeptides, and to the uses of such polynucleotides and polypeptides. More specifically, the invention relates to the phosphomevalonate kinase (PMK) gene (ERG8 gene) from *Candida Albicans* (*C. albicans*), to methods for its expression yielding phosphomevalonate kinase protein, to hybrid organisms for use in such expression methods, to methods for purification of the protein, to methods and tools for diagnosing *C. albicans* infection and to assays for identifying inhibitors of the enzyme which inhibitors have potential as anti-fungal agents.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 00/03100

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/54 C12N9/12 C12Q1/68 G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS, EMBL, STRAND, MEDLINE, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>TSAY Y H ET AL.: "Cloning and characterization of ERG8, an essential gene of <i>Saccharomyces cerevisiae</i> that encodes phosphomevalonate kinase." MOLECULAR AND CELLULAR BIOLOGY, vol. 11, no. 2, February 1991 (1991-02), pages 620-631, XP000981936 abstract figure 4 page 629, right-hand column, line 39-43</p> <p>--- -/--</p>	1-11, 13-15



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

14 February 2001

Date of mailing of the international search report

26/02/2001

Name and mailing address of the ISA

European Patent Office, P.O. Box 5818 Patentplatz 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

van de Kamp, M

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/03100

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LEES N D ET AL.: "Biochemistry and molecular biology of sterol synthesis in <i>Saccharomyces cerevisiae</i> " CRITICAL REVIEWS IN BIOCHEMISTRY AND MOLECULAR BIOLOGY, vol. 34, no. 1, 18 January 1999 (1999-01-18), pages 33-47, XP000981947 page 34, line 14 -page 35, line 3 page 37, line 46 -page 38, line 7 ---	1-11, 13-15
Y	HARTMAN P G ET AL.: "Inhibitors of ergosterol biosynthesis as antifungal agents" CURRENT PHARMACEUTICAL DESIGN, vol. 3, no. 2, April 1997 (1997-04), pages 177-208, XP000981960 abstract page 178, left-hand column, line 20-37 page 179, left-hand column, line 39 -page 181, left-hand column, line 34 ---	1-11, 13-15
A	JENSEN-PERGAKES K L ET AL: "Sequencing, disruption, and characterization of the <i>Candida albicans</i> sterol methyltransferase (ERG6) gene: drug susceptibility studies in <i>erg6</i> mutants." ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, vol. 42, no. 5, May 1998 (1998-05), pages 1160-1167, XP000981986 abstract ---	1-11
A	KELLY R ET AL.: "Cloning and characterization of the 2,3-oxidosqualene cyclase-coding gene of <i>Candida albicans</i> " GENE, vol. 87, no. 2, 15 March 1990 (1990-03-15), pages 177-183, XP002160270 abstract ---	1-11
A	MANNARELLI B M ET AL.: "Rapid identification of <i>Candida albicans</i> and other human pathogenic yeasts by using short oligonucleotides in a PCR" JOURNAL OF CLINICAL MICROBIOLOGY, vol. 36, no. 6, June 1998 (1998-06), pages 1634-1641, XP002160271 abstract ---	13-15
A	US 5 885 815 A (KAUFMANN CHRISTOPH ET AL) 23 March 1999 (1999-03-23) the whole document -----	1-11

INTERNATIONAL SEARCH REPORT

International Application No. PCT/GB 00 03100

FURTHER INFORMATION CONTINUED FROM PCT/SA/ 210

Continuation of Box I.2

Claims Nos.: 12

Remark: Claim 12 refers to a compound that modulates the activity of PMK without giving a true technical characterisation. Moreover, no specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/GB 00/03100

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5885815 A	23-03-1999	NONE	